Gynecologic Oncology 141 (2016) 95–100

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

Molecular classification of high grade endometrioid and clear cell ovarian cancer using TCGA gene expression signatures

Boris Winterhoff [a,1], Habib Hamidi [b,1], Chen Wang [c], Kimberly R. Kalli [c], Brooke L. Fridley [d], Judy Dering [b], Hsiao-Wang Chen [b], William A. Cliby [e], He-Jing Wang [f], Sean Dowdy [e], Bobbie S. Gostout [e], Gary L. Keeney [e], Ellen L. Goode [c], Gottfried E. Konecny [b,*]

[a] Department of Obstetrics/Gynecology & Women's Health, University of Minnesota, Minneapolis, MN, United States
[b] Division of Hematology-Oncology, Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States
[c] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States
[d] Department of Biostatistics, University of Kansas, Kansas City, KS, United States
[e] Department of Gynecologic Surgery, Mayo Clinic, Rochester, MN, United States
[f] Department of Biostatistics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States

HIGHLIGHTS

- TCGA gene signatures are present in high grade case of rare histology.
- Early stage high grade cases of rare histology cluster separately from advanced stages.
- TCGA signatures may help to stratify patients with rare high grade ovarian cancer.

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form 15 January 2016
Accepted 21 February 2016

Keywords:
Ovarian cancer
Molecular subtypes
Endometrioid
Clear cell and high grade serous histologies

ABSTRACT

Background. It is unclear whether the transcriptional subtypes of high grade serous ovarian cancer (HGSOC) apply to high grade clear cell (HGCCOC) or high grade endometrioid ovarian cancer (HGEOC). We aim to delineate transcriptional profiles of HGCCOCs and HGEOCs.

Methods. We used Agilent microarrays to determine gene expression profiles of 276 well annotated ovarian cancers (OCs) including 37 HGCCOCs and 66 HGEOCs. We excluded low grade OCs as these are known to be distinct molecular entities. We applied the prespecified TCGA and CLOVAR gene signatures using consensus non-negative matrix factorization (NMF).

Results. We confirm the presence of four TCGA transcriptional subtypes and their significant prognostic relevance ($p < 0.001$) across all three histological subtypes (HGSOC, HGCCOC and HGEOCs). However, we also demonstrate that 22/37 (59%) HGCCOCs and 30/67 (45%) HGEOCs form 2 additional separate clusters with distinct gene signatures. Importantly, of the HGCCOC and HGEOCs that clustered separately 62% and 65% were early stage (FIGO I/II), respectively. These finding were confirmed using the reduced CLOVAR gene set for classification where most early stage HGCCOCs and HGEOCs formed a distinct cluster of their own. When restricting the analysis to the four TCGA signatures (ssGSEA or NMF with CLOVAR genes) most early stage HGCCOCs and HGEOC were assigned to the differentiated subtype.

Conclusions. Using transcriptional profiling the current study suggests that HGCCOCs and HGEOCs of advanced stage group together with HGSOCs. However, HGCCOCs and HGEOCs of early disease stages may have distinct transcriptional signatures similar to those seen in their low grade counterparts.

1. Introduction

Microarray-based gene expression studies demonstrate that ovarian cancer (OC) is both a clinically diverse and molecularly heterogeneous disease, comprising subtypes with distinct gene expression patterns that are each associated with statistically significant different clinical outcomes. A gene expression analysis of high-grade serous and endometrioid OCs as part of the Australian Ovarian Cancer Study identified distinct

* Corresponding author at: David Geffen School of Medicine, University of California Los Angeles, 2825 Santa Monica Blvd., Suite 200, Santa Monica, CA 90404-2429.
E-mail address: gkonecny@mednet.ucla.edu (G.E. Konecny).
[1] Contributed equally.

http://dx.doi.org/10.1016/j.ygyno.2016.02.023
0090-8258/

molecular subtypes that have been designated with neutral descriptors (C1, C2, C4, and C5) [1]. The four molecular subtypes were validated in 489 high grade serous ovarian cancer (HGSOC) cases using 1500 intrinsically variable genes for consensus non-negative matrix factorization (NMF) clustering and were termed immunoreactive, differentiated, proliferative and mesenchymal on the basis of gene expression in the clusters [2]. These four molecular subtypes have been independently validated and have been shown to be of independent prognostic relevance [3]. Using the TCGA ovarian cancer data set, Verhaak et al. recently confirmed the four molecular subtypes of high grade serous ovarian cancer (HGSOC) using a reduced subtype gene expression signature, named "Classification of Ovarian Cancer" (CLOVAR) [4]. This reduced CLOVAR gene signature is composed of a 100 genes capable of predicting the ovarian cancer subtypes [4]. Validation studies in independent data sets demonstrated that the CLOVAR signature classifies HGSOC with small error rates, making implementation using medium-throughput expression profiling platforms feasible [4].

The main objective of a molecular classification of OC into subtypes with distinct gene expression patterns is to develop robust biomarker signatures that will allow clinicians to identify women likely to benefit from a given therapy. These evolving subgroups are thought to have distinct biologic features that can translate into different therapeutic implications. Epithelial ovarian cancer is a heterogeneous disease consisting of tumors with different histology and grade. The most common OC types are the serous tumors followed by endometrioid and clear-cell cancers which represent 50%–60%, 25% and 4% of all ovarian tumors, respectively [5]. Importantly, however, the evolving molecular classification using the four main subtype signatures have almost exclusively been studied and applied to HGSOC [2,3,4]. Although some early gene expression studies have included endometrioid and clear cell ovarian cancers [6–10] these studies were limited by their small sample size and the use of early generation microarrays. Nevertheless these studies did suggest that clear cell and endometrioid ovarian cancers may be distinguished from serous ovarian cancers based on their gene expression profiles [6–10]. However, many of these early studies included well differentiated tumors (G1) known to be distinct molecular entities [11]. To date it is unclear if the evolving signatures which have been used to successfully classify HGSOC into four molecular subtypes could also be used to classify these less common epithelial ovarian cancer histologies. Although clear cell carcinomas and endometrioid carcinomas have been previously shown to be in part driven by pathways distinct from those driving progression of HGSOC we wanted to investigate whether high grade clear cell ovarian cancers (HGCCOCs) or high grade endometrioid ovarian cancers (HGEOCs) may nevertheless in part share gene signatures that have been described in HGSOCs. For instance, we thought it would be important to know if gene signatures characterizing an immunoreactive or mesenchymal subtype can also be found in HGEOCs or HGCCOCs because the evolving molecular signatures are becoming increasingly clinically relevant. In the present study we, therefore, examined the transcriptional profiles of 276 ovarian cancer cases including 37 HGCCOCs, 66 HGEOCs and 173 previously published HGSOCs using Agilent Whole Human Genome 4x44K Expression Arrays [3]. All low grade tumors were excluded from this study as they are known to represent distinct biologic entities [11]. We applied the pre-specified TCGA gene expression signatures and the reduced CLOVAR gene signatures to this cohort of 276 well annotated OCs from Mayo Clinic. Moreover, we also performed single sample gene set enrichment analysis (ssGSEA) which calculates separate enrichment scores for each sample and allows the assignment to the nearest TGCA subgroup.

2. Materials and methods

2.1. Patient cohort

Fresh frozen tumors were collected from a series of 276 consecutive women with high grade serous, clear cell and endometrioid ovarian, primary peritoneal or fallopian tube cancer who underwent surgery by a gynecologic oncologist at Mayo Clinic between 1994 and 2005. All patients signed an Institutional Review Board approved consent for bio-banking, clinical data extraction, and molecular analysis. Clinical data were abstracted from medical records and tumor registry. Thirteen patients (7.5%) were included in the TCGA study.

2.2. Sample processing and gene expression profiling

Samples were collected during surgery, snap frozen within 30 min, and stored at −80 °C until RNA extraction. Samples were reviewed by a pathologist specialized in gynecologic oncology (G.K.) and selected to have >70% tumor cell content. RNA was isolated using RNeasy (Qiagen Inc., Valencia, CA) and quantified using a Nanodrop Spectrophotomer (Agilent Technologies, Santa Clara, CA). Gene expression profiles were established using Agilent Whole Human Genome 4x44K Expression Arrays. Total RNA (750 ng) with RNA Integrity Number > 8.0 was labeled with cyanine 5-CTP or cyanine 3-CTP using the Low RNA Input Fluorescent Linear Amplification Kit (Agilent Technologies), purified on RNeasy Mini columns (Qiagen Inc.), and hybridized to expression arrays (using a mixed reference containing equal amounts of each of 106 ovarian tumor samples). Slides were scanned using the Agilent 2565BA Scanner and data were exported by the Agilent Feature Extraction Software (version 7.5.1) into Rosetta Resolver (Rosetta Inpharmatics LLC, Cambridge, MA). Log ratios of signal from individual tumor to signal from the reference mix were used for analysis. The data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number: GSE73614.

2.3. Non-negative matrix factorization (NMF) and class prediction

Molecular classification was determined blinded to demographic and clinical information. To identify genes associated with TCGA subtypes, we analyzed expression and molecular subtype data of TCGA cases (https://tcga-data.nci.nih.gov/docs/publications/ov_2011/). Next, we mapped these signature genes to corresponding Agilent probe-set IDs. We selected 1844 probes matching the TCGA signature gene set. Subclasses were computed by reducing the dimensionality of the expression data from thousands of genes to a few metagenes using a consensus NMF clustering method [12]. This method computes multiple k-factor factorization decompositions of the expression matrix and evaluates the stability of the solutions using a cophenetic coefficient. The same analysis was repeated using 161 probes representing 100 genes that represented the CLOVAR subtype signature derived by Verhaak and colleagues [4].

2.4. Gene set enrichment analysis

Gene set activation scores for each of the subtype expression signatures were generated using single sample Gene Set Enrichment Analysis (ssGSEA) [13] using Bioconductor package GSVA downloadable at http://www.bioconductor.org [14]. Raw enrichment scores were expressed as relative z-scores. Subtype assignment of each tumor sample was determined using a z-score cut-off 0.6.

2.5. Statistical analysis of molecular subtypes and patient outcome

Subgroup assignments were compared by use of the chi-square test. Overall survival is depicted according to the method of Kaplan and Meier, and the curves were compared with use of the log rank test. All statistical tests were two-sided.

3. Results

We used Agilent microarrays to determine gene expression profiles of 276 well annotated OCs including 37 HGCCOCs and 66 HGEOCs and